

# SELEKSI FITUR MENGGUNAKAN EKSTRAKSI FITUR BENTUK, WARNA, DAN TEKSTUR DALAM SISTEM TEMU KEMBALI CITRA DAUN

Yuita Arum Sari<sup>1)</sup>, Ratih Kartika Dewi<sup>2)</sup>, Chastine Fatichah<sup>3)</sup>

<sup>1, 2, 3)</sup> Jurusan Teknik Informatika, Fakultas Teknologi Informasi

Institut Teknologi Sepuluh Nopember, Surabaya

Email : yuita.sari12@mhs.if.its.ac.id, ratihkartikadewi@yahoo.com, chastine@cs.its.ac.id

## ABSTRAK

*Fitur yang digunakan untuk mengenali jenis daun meliputi bentuk, warna, dan tekstur. Tidak semua jenis fitur perlu digunakan untuk melakukan komputasi hasil ekstraksi, namun perlu diseleksi beberapa fitur yang paling berpengaruh dalam sistem temu kembali citra daun. Teknik seleksi fitur Correlation based Feature Selection (CFS) digunakan untuk melakukan pemilihan fitur berdasarkan korelasi antar fitur, sehingga dapat meningkatkan performa dari sistem temu kembali citra daun. Jenis seleksi fitur yang digunakan diantaranya menggunakan CFS, CFS dengan Genetic Search (GS), dan chi square. Analisis keterkaitan korelasi antar fitur melalui seleksi fitur juga dikombinasikan dengan penggunaan kedekatan dalam menghitung similaritas pada sistem temu kembali. Penggunaan kedekatan dengan Lp norm, manhattan, euclidean, cosine, dan mahalanobis. Hasil penelitian ini menunjukkan nilai temu kembali paling tinggi ketika menggunakan seleksi fitur CFS dengan pengukuran kedekatan mahalanobis.*

**Kata Kunci:** *Correlation based Feature Selection, Genetic Search, Chi Square, proximity, sistem temu kembali.*

## 1 PENDAHULUAN

Objek daun merupakan salah satu bentuk objek yang memiliki fitur lengkap dari segi bentuk, warna, dan tekstur. Penelitian mengenai pengenalan jenis daun berdasarkan fitur-fitur tersebut telah banyak dilakukan para peneliti.

Penelitian [1] merupakan ringkasan ulang mengenai identifikasi spesies tanaman dari citra digital yang mengusulkan 4 macam fitur untuk mengidentifikasi spesies tanaman, yaitu bentuk daun, pembuluh daun, margin daun (bentuk gerigi pada daun terluar dan tekstur daun. Penelitian ini tidak menggunakan fitur warna.

Penelitian [2] menggunakan fitur warna, bentuk dan tekstur untuk temu kembali gambar daun. Fitur warna dianalisis menggunakan *co-occurent matrix*. Fitur bentuk menggunakan fitur lokal dan fitur global. Fitur tekstur menggunakan Gabor wavelet.

Penelitian ini menggunakan tiga jenis fitur yang melibatkan bentuk, warna, dan tekstur. Ekstraksi fitur bentuk yang digunakan meliputi *slimness, form factor/roundness, rectangularity, narrow factor*, rasio antara keliling dan diameter, dan rasio antara keliling dan panjang serta lebar daun.

Pada ekstraksi fitur warna, gambar dihitung menggunakan model warna RGB. Masing-masing *channel* warna R, G, dan B didapatkan nilai *mean, skewness* dan *kurtosis*. Ekstraksi fitur tekstur menggunakan *gliding box lacunarity* yang

merupakan salah satu teknik untuk mengenali jenis tekstur yang terdapat dalam sebuah gambar [3].

Penelitian [4] menggunakan fitur warna, bentuk dan tekstur untuk analisis gambar daun. Fitur bentuk dianalisis menggunakan *rectangularity, circularity, sphericity, eccentricity, axis ratio, diameter, complexity* dan *perimeter*. Fitur tekstur menggunakan *energy, entropy, correlation, contrast, homogeneity, sum of squares, inverse different moment*, dan *angular second moment*. Fitur warna menggunakan *mean*, standar deviasi, *skewness* dan *kurtosis*. Dari fitur-fitur tersebut kemudian dilakukan seleksi fitur. Seleksi fitur merupakan proses untuk menemukan hasil fitur yang berkorelasi satu sama lain tanpa harus menggunakan keseluruhan hasil dari hasil ekstraksi fitur.

Seleksi fitur pada penelitian [4] menggunakan *Genetic Algorithm* dan Kernel PCA dengan tujuan memilih subset terbaik. Penggabungan GA dan KPCA tersebut digunakan untuk melakukan klasifikasi dengan menggunakan SVM. Hasil penelitian tersebut dengan menggunakan seleksi fitur lebih efektif secara komputasi, dan dapat meningkatkan akurasi.

Proses seleksi fitur juga dapat digunakan untuk mengenali jenis daun melalui sistem temu kembali. Sistem temu kembali adalah salah satu teknik yang digunakan untuk mencocokkan gambar yang berasal dari *query* dengan yang terdapat pada *database* berdasarkan pengukuran kedekatan. Sistem temu kembali informasi gambar dapat digunakan dalam

pengenalan pola daun. Pengenalan pola daun dapat dilakukan dengan menggunakan ekstraksi fitur pada daun melalui fitur bentuk, warna, dan tekstur.

Penelitian [5] melakukan sebuah eksperimen terhadap hasil uji coba sistem temu kembali dengan objek daun menggunakan metode Nearest Neighbour (NN). Fitur yang digunakan fokus pada fitur bentuk dengan dilengkapi fitur pembuluh pada daun. Penelitian tersebut menggunakan sistem pembobotan dengan matriks dan menggunakan derajat similaritas antara dua daun.

Berdasarkan paparan diatas, pada makalah ini akan dilakukan eksplorasi menggunakan teknik seleksi fitur pada ekstraksi fitur bentuk, warna, dan daun untuk meningkatkan performa sistem temu kembali.

## 2 EKSTRAKSI FITUR

Ekstraksi fitur citra daun berupa fitur bentuk, warna dan tekstur.

### 2.1 Ekstraksi Fitur Bentuk Daun

Jenis ekstraksi fitur daun yang digunakan terdiri dari enam jenis fitur, diantaranya adalah [2] :

a) *Slimness*

*Slimness* merupakan rasio perbandingan antara panjang daun dan lebar daun [6].

$$\text{slimness} = \frac{L_p}{W_p} \quad (1)$$

dimana  $L_p$  adalah panjang daun, dan  $W_p$  adalah lebar daun.

b) *Form Factor / Roundness*

$$\text{roundness} = \frac{4\pi A}{P^2} \quad (2)$$

$$\text{roundness} = \frac{4\pi A}{P^2} \quad (3)$$

dimana  $A$  adalah wilayah/ luas area daun dan  $P$  adalah keliling daun.

c) *Rectangularity*

*Rectangularity* adalah teknik untuk menggambarkan kemiripan bentuk daun dengan bentuk kotak [7].

$$\text{rectangularity} = \frac{L_p W_p}{A} \quad (4)$$

$$\text{rectangularity} = \frac{L_p W_p}{A} \quad (5)$$

dimana  $L_p$  adalah panjang daun,  $W_p$  adalah lebar daun, dan  $A$  adalah wilayah/ luas area daun.

d) *Narrow factor*

*Narrow factor* didefinisikan sebagai rasio dari diameter  $D$  dan panjang dari daun  $L_p$ .

$$\text{narrow factor} = \frac{D}{L_p} \quad (6)$$

e) Rasio keliling dan diameter

Perbandingan antara keliling ( $P$ ) dan diameter ( $D$ ) digunakan sebagai salah satu fitur bentuk daun.

$$\text{ratio} = \frac{P}{D} \quad (7)$$

f) Rasio perimeter dengan panjang dan lebar

Perbandingan antara keliling ( $P$ ) dan panjang ( $L_p$ ) serta lebar daun ( $W_p$ ) digunakan sebagai salah satu fitur bentuk daun.

$$\text{ratio} = \frac{P}{(L_p + W_p)} \quad (8)$$

### 2.2 Ekstraksi Fitur Warna Daun

Ekstraksi fitur warna pada daun terdiri dari mean, *skewness* dan *kurtosis*:

a) *Mean*

*Mean* merupakan rata-rata nilai piksel ( $P_{ij}$ ) pada masing-masing *channel* R, G dan B.

$$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N P_{ij} \quad (9)$$

dimana  $M$  dan  $N$  secara berturut-turut adalah nilai panjang dan lebar piksel pada gambar.

b) *Skewness* dan *kurtosis*

*Skewness* dan *kurtosis* merupakan rata-rata nilai piksel ( $P_{ij}$ ) dikurangi *mean* ( $\mu$ ) kemudian dipangkatkan dengan  $n$  pada masing-masing *channel* R, G dan B.

$$\sigma = \frac{\sum_{i=1}^M \sum_{j=1}^N (P_{ij} - \mu)^n}{MN \sigma^n} \quad (10)$$

dimana untuk *skewness* nilai  $n$  adalah 3, dan untuk *kurtosis* nilai  $n$  adalah 4.

*Skewness* adalah derajat ketidaksimetrisan suatu distribusi. Apabila *skewness* = 0 maka dikatakan simetris. *Kurtosis* adalah derajat keruncingan suatu distribusi (biasanya diukur relatif terhadap distribusi normal). Dalam penelitian ini distribusi yang dimaksud adalah warna. Secara perhitungan, *skewness* adalah momen ketiga terhadap *mean* dan *kurtosis* adalah momen keempat terhadap *mean* [8]. Oleh karena itu, pada rumus di atas  $n$  bernilai 3 untuk *kurtosis* dan bernilai 4 untuk *skewness*. Histogram adalah teknik grafis yang efektif untuk menunjukkan *skewness* dan *kurtosis* dari sebuah *dataset*.

### 2.3 Ekstraksi Fitur Tekstur Daun

Ekstraksi fitur tekstur dari gambar dilakukan dengan *gliding box lacunarity* untuk estimasi dimensi fraktal. Hal ini disebabkan tekstur dapat terkarakterisasi dari pendefinisian pengukuran lacunarity. *Lacunarity Gliding Box* (LGB)

merupakan penghitungan dimensi fraktal dari B (lakuna) yang masukannya berupa titik data biner 0 dan 1, 0 menyatakan lubang atau lakuna [3].

$$LGB(r) = \frac{B(r) \sum_{i=1}^{B(r)} p(i,r)^2}{\sum_{i=1}^{B(r)} p(i,r)} \quad (11)$$

dimana  $B(r)$  adalah kotak dengan ukuran  $r$ ,  $r$  merupakan ukuran yang spesifik dan  $p$  adalah banyaknya titik dalam kotak tersebut. LGB merupakan rasio momen kedua dan momen pertama dari *gliding box*.

### 3 METODE SELEKSI ATRIBUT

Metode seleksi atribut meliputi *correlation based feature selection (CFS)*, *CFS* dengan *Genetic Search (GS)* dan *Chi Square*.

#### 3.1 Correlation based Feature Selection (CFS)

Penelitian mengenai pemilihan atribut secara otomatis berkembang dalam ranah pembelajaran mesin. Salah satu metode untuk pemilihan atribut secara otomatis adalah CFS (*Correlation based Feature Selection*). CFS mengidentifikasi fitur yang relevan, artinya tidak ada ketergantungan yang kuat dengan fitur lainnya. Seleksi fitur dapat memperbaiki performa akurasi atau paling tidak menghasilkan akurasi yang sama dengan metode tanpa seleksi fitur. Metode yang digunakan dalam CFS adalah menentukan pasangan fitur dan menentukan bobot fitur dengan algoritma RELIEF [9]. Seleksi atribut dengan CFS menggunakan algoritma sebagai berikut:

- Penghitungan nilai korelasi antar fitur dan fitur-kelas dengan Relief.
- Merit menyatakan himpunan fitur yang berkorelasi tinggi dengan kelas, tetapi tidak berkorelasi dengan masing-masing fitur lain. *rcf* adalah nilai tengah dari korelasi fitur-kelas dan *rff* adalah nilai tengah dari korelasi fitur-fitur.  $K$  adalah fitur dan merit  $k$  merupakan nilai merit (kontribusi sebuah fitur dalam menentukan hasil klasifikasi) dari sebuah fitur

$$Merit_k = \frac{krcf}{\sqrt{k + k(k+1)rff}} \quad (12)$$

- CFS dapat dihitung dari pengembangan merit, yaitu dengan nilai merit terbaik (nilainya paling besar).

##### 3.1.1 CFS dengan GS

Pencarian genetik merepresentasikan pencarian dengan algoritma genetik. Pencarian genetik diterapkan pada CFS untuk mendapatkan pemilihan fitur yang paling optimal. Algoritma genetik memiliki alur sebagai berikut [10]:

- Inisialisasi populasi, dilakukan secara *random*.

- Seleksi kromosom berdasarkan evaluasi *Fitness*. Teknik seleksi yang dipakai dalam kebanyakan aplikasi adalah seleksi *roulette-wheel*.
- Kromosom yang terseleksi dilakukan pindah silang dan mutasi sehingga terbentuk populasi yang baru.
- Populasi lama tergantikan dengan populasi baru.
- Proses evolusi akan terus berlangsung sampai kriteria penghentian tercapai.

#### 3.2 Chi-Square

Chi square merupakan metode yang banyak digunakan. Chi square mengevaluasi nilai dari statistik chi-square untuk masing-masing atribut dikaitkan dengan kelas. Hipotesis  $H_0$  merupakan asumsi yang menyatakan 2 fitur saling tidak berkorelasi dan diuji dengan:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (13)$$

$O_{ij}$  adalah frekuensi yang diobservasi dan  $E_{ij}$  adalah frekuensi yang diperkirakan (secara teoretis). Makin besar nilai chi square maka makin besar penentangan terhadap hipotesis  $H_0$  [11].

### 4 PENGUKURAN KEDEKATAN

Pengukuran kedekatan meliputi *minkowski*, *manhattan distance*, *Euclidean*, *mahalanobis*, dan *cosine similarity*.

#### 4.1 Minkowski dengan $L_p$ -Norm

Pendekatan yang digunakan dalam penelitian ini menggunakan *minkowski distance* atau pengukuran jarak dengan menggunakan  $L_p$ -Norm, dimana nilai input berupa  $p$ . Nilai  $p$  adalah nilai bilangan bulat positif. Jika nilai  $p=2$ , maka disebut juga dengan  $L_2$ -norm atau *euclidean distance*. Persamaan mengenai *minkowski distance* dapat dijelaskan pada persamaan

$$d_{x1} = \sqrt[p]{\sum_{i=1}^n |x_{s1} - x_{t1}|^p} \quad (14)$$

dimana,  $x_s$  dan  $x_t$  merupakan dua buah vektor yang dicari kedekatannya,  $p$  merupakan masukan parameter sistem dengan nilai  $p$  berupa bilangan bulat positif, dan  $n$  merupakan banyaknya data yang terbentuk berupa matriks [12].

#### 4.2 Manhattan Distance

*Manhattan distance* merupakan pengembangan dari  $L_p$  norm dengan  $p=1$  [13].

#### 4.3 Euclidean

*Euclidean distance* merupakan pengembangan dari  $L_p$  norm dengan  $p=2$  [13].

#### 4.4 Mahalanobis

Jarak mahalanobis menghitung jarak antara  $x$  dan  $y$  dengan menyertakan *covariance matrix* ( $C$ ) [14]

$$d_{x,y}^2 = (X_s - Y_t)C^{-1}(X_s - Y_t)^T \quad (15)$$

dimana  $C$  adalah *covariance matrix*,  $X_s$  adalah vektor pertama dan  $Y_t$  adalah vektor kedua, dan  $d_{x,y}^2$  adalah kuadrat dari jarak mahalanobis sehingga hasil jarak mahalanobis adalah akar dari persamaan tersebut. Jarak antara  $X_s$  dan  $Y_t$  merepresentasikan jarak yang dimaksud dalam mahalanobis.

#### 4.5 Cosine similarity

Dua buah vektor  $X=(x_1, x_2, \dots, x_n)$  dan  $Y=(y_1, y_2, \dots, y_n)$  maka  $\cos$  adalah *cosine* dari sudut antara  $X$  dan  $Y$  pada  $n$  dimensi [13].

$$VCAD(X, Y) = \frac{X \cdot Y}{\|X\| \|Y\|} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \sqrt{\sum_i y_i^2}} \quad (16)$$

*Cosine similarity* (VCAD/ Vector cosine angle distance) merupakan pengukuran similaritas antara 2 vektor melalui *cosine* sudut antara vector  $X$  dan  $Y$ .

## 5 SKENARIO UJI COBA

Pada subbab ini dibahas mengenai analisis data dan langkah-langkah yang dilakukan dalam penelitian. Perangkat bantu yang digunakan untuk melakukan skenario pengujian adalah MATLAB dan WEKA. Ekstraksi fitur bentuk, warna, dan tekstur digunakan MATLAB, sedangkan proses untuk mengetahui adanya korelasi antara fitur dengan melakukan seleksi fitur digunakan WEKA. Fitur yang telah diseleksi akan diproses kembali untuk mendapatkan teknik CBIR citra daun dengan menggunakan MATLAB.

### 5.1 Analisis Data

Data yang berupa macam-macam citra daun diambil dari Flavia (<http://flavia.sourceforge.net/>). Data gambar daun yang digunakan 300 citra daun yang terbagi atas 10 kelas. Data akan dipisahkan menjadi dua bagian yaitu data *corpus*, dan data *query*. Data *corpus* merupakan data yang digunakan untuk pelatihan sampai mendapatkan ciri yang dibutuhkan data untuk dimasukkan ke dalam *database*. Sedangkan data *query* merupakan data yang digunakan sebagai masukan yang akan diekstraksi fitur-fiturnya, kemudian dicocokkan dengan *database* untuk mendapatkan hasil *ranking* kemiripan antara *database* dengan data masukan. Data *query* yang digunakan sebanyak 30 data dengan 10 kelas yang berbeda (Gambar 1).

### 5.2 Rancangan Sistem

Langkah-langkah penelitian diawali dengan proses *preprocessing* pada gambar citra daun. Teknik *preprocessing* adalah rangkaian ekstraksi fitur yang

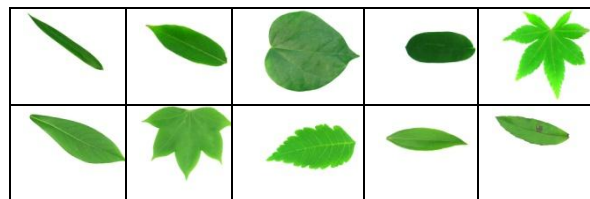
digunakan untuk mengekstrak informasi yang terdapat pada citra daun. Langkah awal dalam *preprocessing* adalah mengenali objek citra daun dengan latar belakang yang digunakan, konversi yang digunakan adalah hitam untuk latar belakang gambar dan gambar putih untuk objek citra daun yang diolah informasinya. Gambar 2 menunjukkan perubahan dari citra daun asli ke dalam proses segmentasi.

Masing-masing data *corpus* citra dan *query* diekstraksi menggunakan fitur bentuk, warna, dan tekstur, sebagaimana telah dijelaskan dalam bab 2. Ekstraksi fitur digunakan untuk mendapatkan hasil keseluruhan fitur yang kemudian akan digunakan seleksi fitur untuk mengetahui keterkaitan antar fitur yang paling cocok. Metode seleksi fitur yang digunakan dijelaskan pada bab 3. Atribut hasil seleksi dari masing-masing metode diukur kedekatannya dengan pencocokan menggunakan berbagai macam metode pengukuran jarak atau similaritas, sebagaimana dijelaskan pada bab 4. Hasil dari proses pencocokan adalah temu kembali citra daun yang mirip antara satu dengan lainnya. Gambar 3 menunjukkan alur sistem secara umum.

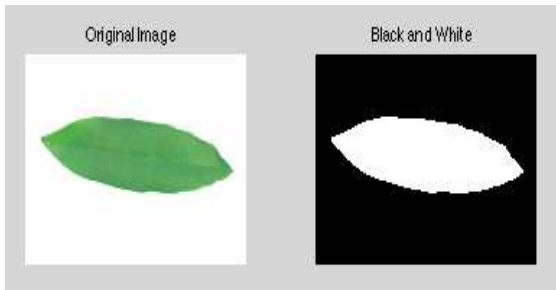
### 5.3 Langkah Uji Coba

Evaluasi yang digunakan untuk melakukan hasil dan analisis digunakan evaluasi secara *ranked retrieval* dan *unranked retrieval*. Proses evaluasi dengan menggunakan *ranked retrieval* digunakan akurasi, *precision*, *recall*, dan *F-measure*. Evaluasi secara *ranked retrieval* digunakan MAP. Pengujian dilakukan pada data *query* sebanyak 30 citra daun dengan *top-30*. Pemberian *top-30* karena diasumsikan bahwa dalam satu kelas atau satu jenis daun mengandung masing-masing 30 citra di dalam *database*.

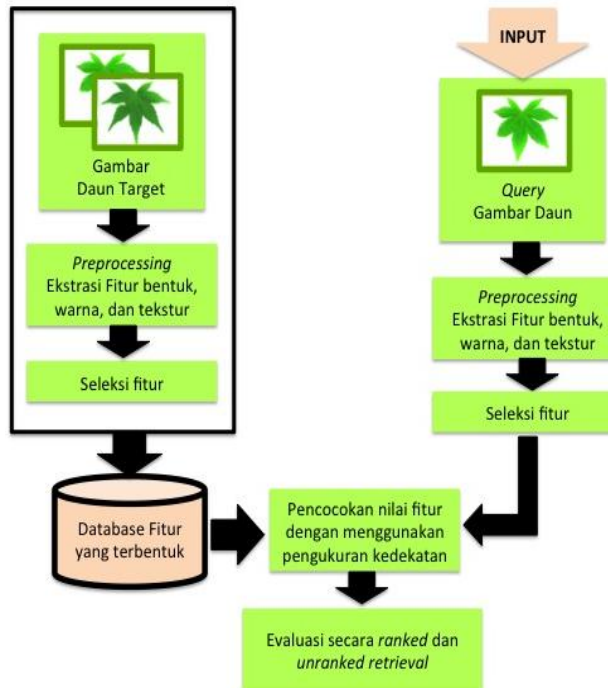
Pengujian dilakukan dengan menggunakan fungsi kedekatan yang berbeda-beda, diantaranya adalah  $L_p$ -norm, dengan  $p = 1$  merupakan jarak Manhattan dan  $p = 2$  merupakan jarak Euclidean. Uji coba juga dilakukan pada  $p = 0.5$ ,  $p = 10$ , dan  $p = 100$ . Selain menggunakan jarak  $L_p$ -norm digunakan pula jarak *cosine* dan mahalanobis. Evaluasi ditinjau dari hasil sebelum mengalami seleksi fitur, dan telah mengalami seleksi fitur.



**Gambar 1.** Jenis daun yang merepresentasikan 10 kelas yang berbeda



Gambar 2. Segmentasi citra daun



Gambar 3. Alur sistem secara umum

## 6 HASIL UJI COBA

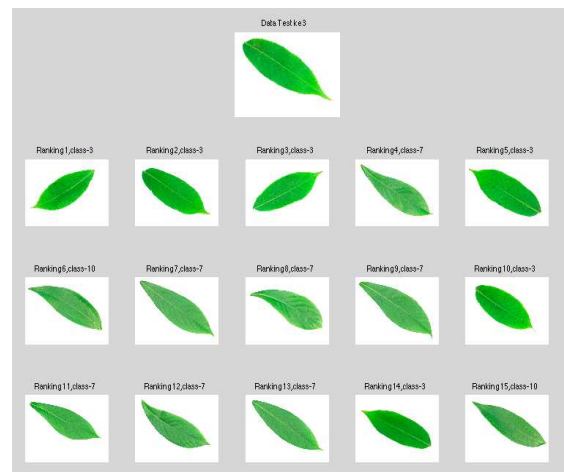
Pengambilan citra digunakan 30 nilai teratas dengan ukuran jarak antara citra *query* dengan citra dalam *database* yang semakin kecil, sehingga menunjukkan similaritas yang sesuai. Contoh hasil dari sistem CBIR ditampilkan pada Gambar 4. Hasil dari proses ekstraksi fitur menghasilkan 16 fitur utama yang digunakan untuk mengenali ciri yang saling mirip dalam proses temu kembali citra gambar daun. Tabel 1 menunjukkan hasil evaluasi sistem temu kembali citra daun sebelum mengalami ekstraksi fitur.

Penggunaan jarak mahalanobis memiliki nilai akurasi, *precision*, *recall*, dan MAP paling bagus dibandingkan dengan penggunaan metode jarak yang lain. Pengukuran jarak dengan menggunakan *cosine* terbukti kurang efektif dalam menentukan hasil informasi dari fitur-fitur yang diharapkan oleh *query*. Secara umum, hasil evaluasi jarak *lp-norm*

menunjukkan bahwa semakin kecil nilai parameter  $p$  semakin bagus dalam meningkatkan hasil evaluasi.

Fitur yang terlalu banyak mengakibatkan pemberian informasi antar fitur menjadi terlalu renggang. Seleksi fitur digunakan untuk menentukan korelasi antar fitur yang memiliki informasi yang saling berkaitan lebih banyak dengan menggunakan sistem *ranking*. Hasil seleksi fitur terdapat pada Tabel 2. Seleksi fitur dengan menggunakan metode CFS menghasilkan 8 fitur, dimana fitur tekstur, tidak terseleksi dengan menggunakan metode CFS. Sehingga, dalam metode CFS hanya digunakan ekstraksi unsur fitur pokok bentuk dan warna. Hasil seleksi fitur dengan menggunakan kombinasi CFS dan *Genetic Search* (CFS+GS) menghasilkan 8 fitur, dimana kandungan fitur-fitur yang terseleksi melibatkan unsur bentuk, warna dan tekstur. Seleksi fitur dengan *chi-square* menghasilkan *ranking* secara terurut, digunakan fitur yang mengandung bentuk, warna, dan tekstur, sehingga terdapat 11 fitur yang terseleksi.

Ketiga metode seleksi fitur tersebut, menghasilkan hasil seleksi fitur yang saling beririsan satu sama lain, diantaranya fitur *slimness*, *roundness*, *rectangularity*, rasio perbandingan perimeter dengan panjang dan lebar. Keseluruhan fitur yang saling beririsan satu dengan yang lain hanya dari fitur bentuk. Seleksi fitur warna antar metode dapat berbeda-beda, hal ini disebabkan karena warna daun yang digunakan dalam *dataset* berwarna hijau.



Gambar 4. Evaluasi data test 3

Tabel 1. Evaluasi sebelum mengalami seleksi fitur

Jarak	Akurasi	Precision	Recall	F-Measure	MAP
L-100	0,58	0,58	0,29	0,39	0,49
L-10	0,49	0,49	0,25	0,33	0,39
L-0.5	0,58	0,58	0,29	0,39	0,50
L-1 (Manhattan)	0,51	0,51	0,26	0,34	0,43
L-2 (Euclidean)	0,51	0,51	0,25	0,34	0,41



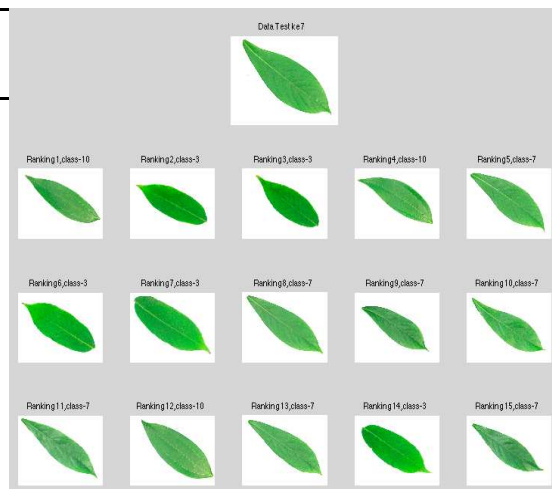
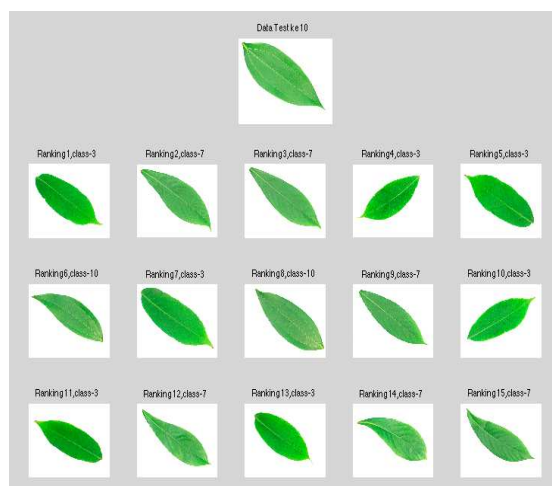
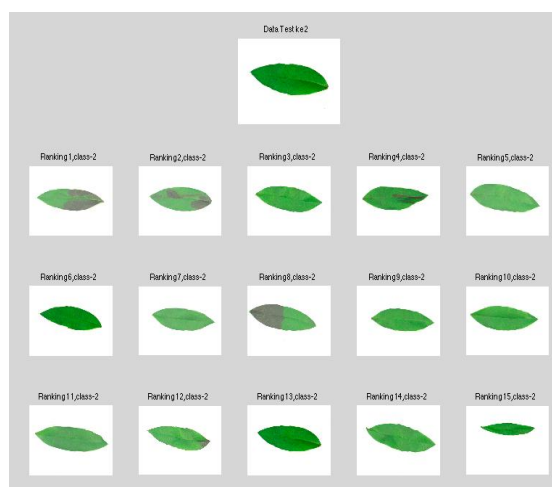
<b>Cosine distance</b>	0,07	0,07	0,03	0,04	0,04
<b>Mahalanobis</b>	0,73	0,73	0,37	0,49	0,67

**Tabel 2.** Hasil fitur yang terbentuk dari seleksi fitur

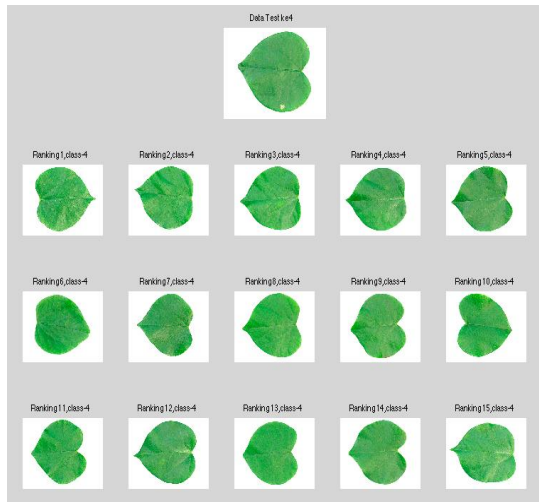
Seleksi Fitur	Fitur yang terseleksi
<b>CFS</b>	<i>Slimness, roundness, rectangularity, rasio perimeter dengan panjang dan lebar, rata-rata warna pada channel G, rata-rata warna pada channel B, kurtosis pada channel B, dan skewness pada channel R.</i>
<b>CFS + GS</b>	<i>Slimness, roundness, rectangularity, rasio keliling dan diameter, rata-rata warna pada channel R, kurtosis pada channel B, skewness pada channel R, dan rata-rata gliding box lacunarity.</i>
<b>Chi-Square</b>	<i>Rectangularity, rasio keliling dan diameter, roundness, slimness, skewness pada channel B, skewness pada channel G, skewness pada channel R, rasio perimeter dengan panjang dan lebar, narrow factor, rata-rata warna pada channel G, dan rata-rata gliding box lacunarity.</i>

Penggunaan fitur *skewness* dan *kurtosis* dalam pemilihan fitur warna menjadi penting, karena dapat meningkatkan performa sistem temu kembali, dibandingkan tanpa menggunakan *skewness* dan *kurtosis*. Fitur tekstur yang digunakan tidak dapat menjadi fitur yang nilai korelasinya tertinggi, karena fitur tekstur, dinilai dari jenis tulang daun yang tidak terlihat mencolok antara citra daun yang satu dengan citra daun yang lain.

Hasil evaluasi menunjukkan adanya dataset yang ambigu antara citra daun yang ada antara kelas 3, kelas 7, dan kelas 10 pada data test (Gambar 5-6). Sehingga akurasi yang dihasilkan ketika menggunakan dataset pada kelas tersebut rendah, karena dinilai bahwa daun yang diteliti mempunyai bentuk, warna, dan tekstur yang mirip. Pada datatest 2 (Gambar 7), sistem dapat mengenali jenis daun pada kelas yang sama dengan skala yang berbeda. Sistem telah dapat mengenali adanya rotasi pada gambar, terbukti pada datatest ke-4 dan ke-6 (Gambar 8-9).

**Gambar 5.** Evaluasi data test 7**Gambar 6.** Evaluasi data test 10**Gambar 7.** Evaluasi data test 2

Hasil evaluasi masing-masing seleksi fitur CFS, CFS+GS, dan *Chi-square* ditunjukkan pada Tabel 3, 4, dan 5 secara berurutan.



Gambar 8. Evaluasi data test 4



Gambar 9. Evaluasi data test 6

Tabel 3. Evaluasi dengan menggunakan CFS

Jarak	Akurasi	Precision	Recall	F-Measure	MAP
L-100	0,580	0,580	0,290	0,387	0,493
L-10	0,500	0,500	0,250	0,333	0,404
L-0.5	0,613	0,613	0,307	0,409	0,554
L-1 (Manhattan)	0,553	0,553	0,277	0,369	0,479
L-2 (Euclidean)	0,500	0,500	0,250	0,333	0,415
Cosine distance	0,053	0,053	0,027	0,036	0,033
Mahalanobis	0,727	0,727	0,363	0,484	0,664

Tabel 4. Evaluasi dengan menggunakan CFS+GS

Jarak	Akurasi	Precision	Recall	F-Measure	MAP
L-100	0,400	0,400	0,200	0,267	0,307
L-10	0,327	0,327	0,163	0,218	0,222
L-0.5	0,560	0,560	0,280	0,373	0,447
L-1 (Manhattan)	0,393	0,393	0,197	0,262	0,294
L-2 (Euclidean)	0,340	0,340	0,170	0,227	0,239
Cosine distance	0,053	0,053	0,027	0,036	0,036
Mahalanobis	0,647	0,647	0,323	0,431	0,563

Tabel 5. Evaluasi dengan menggunakan *Chi-square*

Jarak	Akurasi	Precision	Recall	F-Measure	MAP
L-100	0,507	0,507	0,253	0,338	0,382
L-10	0,460	0,460	0,230	0,307	0,303
L-0.5	0,533	0,533	0,267	0,356	0,436
L-1 (Manhattan)	0,473	0,473	0,237	0,316	0,338
L-2 (Euclidean)	0,453	0,453	0,227	0,302	0,305
Cosine distance	0,060	0,060	0,030	0,040	0,026
Mahalanobis	0,673	0,673	0,337	0,449	0,608

Evaluasi sistem temu kembali citra daun dengan seleksi fitur CFS ditunjukkan nilai akurasi paling tinggi dengan pencocokan menggunakan Mahalanobis. Jika dibandingkan dengan evaluasi sebelum seleksi fitur, secara mayoritas penggunaan seleksi fitur CFS dapat meningkatkan akurasi sistem temu kembali citra daun.

Hasil akurasi ketika menggunakan jarak *cosine* lebih baik tanpa seleksi fitur, daripada menggunakan seleksi fitur. Penggunaan seleksi fitur dengan CFS+GS tidak lebih bagus hasil evaluasinya, dibandingkan dengan menggunakan CFS dan ketika sistem tidak mengalami metode seleksi fitur. Hasil sistem temu kembali citra seleksi fitur *chi-square* menghasilkan akurasi yang paling bagus dengan menggunakan kedekatan Mahalanobis. Jika dibandingkan dengan tanpa menggunakan seleksi fitur, kedekatan *Lp-norm* memiliki hasil yang kurang bagus dibandingkan dengan menggunakan similaritas *cosine* dan jarak mahalanobis.

Kelebihan menggunakan seleksi fitur adalah ketika menggunakan jarak mahalanobis. Hasil evaluasi sistem temu kembali tetap stabil meningkat, dibandingkan tanpa menggunakan seleksi fitur. Mahalanobis cocok digunakan di dalam semua metode seleksi fitur.

Kekurangan seleksi fitur dibandingkan dengan yang tanpa menggunakan seleksi fitur terletak pada penggunaan kedekatan yang berbeda. Penggunaan kedekatan *cosine* menunjukkan hasil evaluasi yang paling buruk untuk sistem temu kembali citra. Secara keseluruhan penggunaan seleksi fitur, kedekatan dengan *Lp-norm* akan semakin buruk jika parameter  $p$  yang digunakan terlalu besar, karena dapat memisahkan lebih jauh informasi antar fitur.

## 7 KESIMPULAN

Penggunaan metode seleksi fitur dengan CFS dan kedekatan mahalanobis paling cocok digunakan dalam sistem temu kembali citra daun. Fitur-fitur dengan korelasi yang paling besar antara satu fitur dengan fitur yang lain dapat meningkatkan performa dari sistem temu kembali citra daun. Fitur warna dan bentuk yang mempunyai pengaruh paling besar

dalam pemilihan fitur dalam *dataset* yang digunakan. Penelitian selanjutnya dapat melakukan optimasi terhadap pemilihan fitur untuk mengenali fitur daun dengan lebih teliti dengan menggunakan metode aturan asosiasi.

## 8 DAFTAR PUSTAKA

- [1] Cope, James S., et al, (2012). "Plant species identification using digital morphometrics: A review." *Expert Systems with Applications* 39,8: 7562-7573.
- [2] Kebapci, Hanife, et all, April (2010). "Plant Image Retrieval Using Color, Shape and Texture Features". *The Computer Journal Advance Access published April 9, 2010*.
- [3] Tolle, Charles R, (2008). "An efficient implementation of the gliding box lacunarity algorithm". *Physica D* 237: 306–315.
- [4] Valiammal, N. dan S.N Geethalaksmi, (2012). "An Optimal Feature Subset Selection for Leaf Analysis". *International Journal of Computer and Communication Engineering* 6.
- [5] Nam, Yunyoung, Eenjun Hwang, and Dongyoon Kim, (2008). "A similarity-based leaf image retrieval scheme: Joining shape and venation features." *Computer Vision and Image Understanding* 110, 2: 245-259.
- [6] Wu, Qingfeng, Changle Zhou, and Chaonan Wang, (2006). "Feature extraction and automatic recognition of plant leaf using artificial neural network". *Advances in Artificial Intelligence* 3.
- [7] Singh, Krishna, Indra Gupta, and Sangeeta Gupta, (2010). "SVM-BDT PNN and Fourier moment technique for classification of leaf shape". *International Journal of Signal Processing, Image Processing and Pattern Recognition* 3, 4: 67-78.
- [8] NIST/SEMATECH, (2013). *E-Handbook of Statistical Methods*. <URL: <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm>>.
- [9] Hall, Mark A, (1999). "Correlation based Feature Selection for Machine Learning". *Tesis of the University of Waikato*. < URL: [www.cs.waikato.ac.nz/~mhall/thesis.pdf](http://www.cs.waikato.ac.nz/~mhall/thesis.pdf)>.
- [10] Zhu, Fangming, Sheng-Uei Guan, (2008). "Cooperative co-evolution of GA-based classifiers based on input decomposition". *Engineering Applications of Artificial Intelligence* 21: 1360-1369.
- [11] Novaković, Jasmina, Perica ŠTRBAC, and Dušan Bulatović, (2011). "Toward optimal feature selection using ranking methods and classification algorithms." *The Yugoslav Journal of Operations Research ISSN: 0354-0243 EISSN: 2334-6043* 21.1.
- [12] Hyun Oh, Jae, and Nojun Kwak, (2013). "Generalization of linear discriminant analysis using Lp-norm". *Pattern Recognition Letters* 34: 679–685.
- [13] Vadivel, A. K. M. S. S. A., A. K. Majumdar, and Shamik Sural, (2003). "Performance comparison of distance metrics in content-based image retrieval applications". *Proc. of Internat. Conf. on Information Technology, Bhubaneswar, India*.
- [14] Kokare, Manesh, B. N. Chatterji, and P. K. Biswas, (2003). "Comparison of similarity metrics for texture image retrieval". *TENCON Conference on convergent technologies for Asia-Pacific region*.